

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Prudent, James R.
Hall, Jeff G.
Lyamichev, Victor I.
- (ii) TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
- (iii) NUMBER OF SEQUENCES: 69
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Medlen & Carroll, LLP
 - (B) STREET: 220 Montgomery Street, Suite 2200
 - (C) CITY: San Francisco
 - (D) STATE: California
 - (E) COUNTRY: United States Of America
 - (F) ZIP: 94104
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/682,853
 - (B) FILING DATE: 12-JUL-1996
- (viii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/599,491
 - (B) FILING DATE: 24-JAN-1996
- (ix) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Ingolia, Diane E.
 - (B) REGISTRATION NUMBER: 40,027
 - (C) REFERENCE/DOCKET NUMBER: FORS-02564
- (x) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (415) 705-8410
 - (B) TELEFAX: (415) 397-8338

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2506 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAGGGGGA	TGCTGCCCT	CTTGAGCCC	AAGGGCCGGG	TCCTCCTGGT	GGACGGCAC	60
CACCTGGCCT	ACCGCACCTT	CCACGCCCTG	AAGGGCCTCA	CCACCAGCCG	GGGGGAGCCG	120
GTGCAGGCGG	TCTACGGCTT	CGCCAAGAGC	CTCCTCAAGG	CCCTCAAGGA	GGACGGGAC	180
GCGGTGATCG	TGGTCTTGA	CGCCAAGGCC	CCCTCCTTCC	GCCACGAGGC	CTACGGGGGG	240
TACAAGGCGG	GCCGGGCCCC	CACGCCGGAG	GACTTCCCC	GGCAACTCGC	CCTCATCAAG	300
GAGCTGGTGG	ACCTCCTGGG	GCTGGCGCCG	CTCGAGGTCC	CGGGCTACGA	GGCGGACGAC	360
GTCCTGGCCA	GCCTGGCAA	GAAGGCGGAA	AAGGAGGGCT	ACGAGGTCCG	CATCCTCACC	420
GCCGACAAAG	ACCTTTACCA	GTCCTTTCC	GACCGCATCC	ACGTCCCTCCA	CCCCGAGGGG	480
TACCTCATCA	CCCCGGCCTG	GCTTTGGAA	AAGTACGGCC	TGAGGCCCGA	CCAGTGGGCC	540
GAATACCGGG	CCCTGACCGG	GGACGAGTCC	GACAACCTTC	CCGGGGTCAA	GGGCATCGGG	600
GAGAACGACGG	CGAGGAAGCT	TCTGGAGGAG	TGGGGGAGCC	TGGAAGGCCCT	CCTCAAGAAC	660
CTGGACCGGC	TGAAGCCCGC	CATCCGGGAG	AAGATCCTGG	CCCACATGGA	CGATCTGAAG	720
CTCTCCTGGG	ACCTGGCAA	GGTGGCGACC	GACCTGCC	TGGAGGTGGA	CTTCGCCAAA	780
AGGCGGGAGC	CCGACCGGGA	GAGGCTTAGG	GCCTTCTGG	AGAGGCTTGA	GTTTGGCAGC	840
CTCCTCCACG	AGTTCGGCCT	TCTGGAAAGC	CCCAAGGCC	TGGAGGAGGC	CCCCTGGCCC	900
CCGCCGGAAG	GGGCCTTCGT	GGGCTTTGTG	CTTCCCAGCA	AGGAGCCAT	GTGGGCCGAT	960
CTTCTGGCCC	TGGCGCCGC	CAGGGGGGGC	CGGGTCCACC	GGGCCCCCGA	GCCTTATAAAA	1020
GCCCTCAGGG	ACCTGAAGGA	GGCGCGGGGG	CTTCTCGCCA	AAGACCTGAG	CGTTCTGGCC	1080
CTGAGGGAAAG	GCCTTGGCCT	CCCGCCCGGC	GACGACCCCA	TGCTCCTCGC	CTACCTCCTG	1140
GACCCCTCCA	ACACCACCCC	CGAGGGGGTG	GCCCGCGCT	ACGGCGGGGA	GTGGACGGAG	1200
GAGGCGGGGG	AGCGGGCCGC	CCTTTCCGAG	AGGCTCTCG	CCAACCTGTG	GGGGAGGCTT	1260
GAGGGGGGAGG	AGAGGCTCCT	TTGGCTTTAC	CGGGAGGTGG	AGAGGCCCCT	TTCCGCTGTC	1320
CTGGCCCACA	TGGAGGCCAC	GGGGGTGCGC	CTGGACGTGG	CCTATCTCAG	GGCCTTGTCC	1380
CTGGAGGTGG	CCGAGGAGAT	CGCCCGCCTC	GAGGCCGAGG	TCTTCCGCCT	GGCCGGCCAC	1440
CCCTTCAACC	TCAACTCCCG	GGACCAGCTG	GAAAGGGTCC	TCTTGACGA	GCTAGGGCTT	1500
CCCGCCATCG	GCAAGACGGA	GAAGACCGGC	AAGCGCTCCA	CCAGCGCCGC	CGTCCTGGAG	1560
GCCCTCCGCG	AGGCCACCC	CATCGTGGAG	AAGATCCTGC	AGTACCGGGA	GCTCACCAAG	1620
CTGAAGAGCA	CCTACATTGA	CCCCTTGCCG	GACCTCATCC	ACCCCAGGAC	GGGCCGCCTC	1680
CACACCCGCT	TCAACCAGAC	GGCCACGGCC	ACGGGCAGGC	TAAGTAGCTC	CGATCCCAAC	1740
CTCCAGAACAA	TCCCCGTCCG	CACCCCGCTT	GGGCAGAGGA	TCCGCCGGGC	CTTCATCGCC	1800

GAGGAGGGGT GGCTATTGGT GGCCCTGGAC TATAGCCAGA TAGAGCTCAG GGTGCTGGCC	1860
CACCTCTCCG GCGACGAGAA CCTGATCCGG GTCTTCAGG AGGGGCGGGA CATCCACACG	1920
GAGACCGCCA GCTGGATGTT CGGCGTCCCC CGGGAGGCCG TGGACCCCCT GATGCGCCGG	1980
GCGGCCAAGA CCATCAACTT CGGGGTCCCTC TACGGCATGT CGGCCAACCG CCTCTCCCAG	2040
GAGCTAGCCA TCCCTTACGA GGAGGCCAG GCCTTCATTG AGCGCTACTT TCAGAGCTTC	2100
CCCAAGGTGC GGGCCTGGAT TGAGAAGACC CTGGAGGAGG GCAGGAGGCG GGGTACGTG	2160
GAGACCTCTC TCGGCCGCCG CCGCTACGTG CCAGACCTAG AGGCCCGGGT GAAGAGCGTG	2220
CGGGAGGCAG CCGAGCGCAT GCCCTTCAAC ATGCCCGTCC AGGGCACCGC CGCCGACCTC	2280
ATGAAGCTGG CTATGGTCAA GCTCTTCCCC AGGCTGGAGG AAATGGGGGC CAGGATGCTC	2340
CTTCAGGTCC ACGACGAGCT GGTCTCGAG GCCCCAAAAG AGAGGGCGGA GGCGTGGCC	2400
CGGCTGGCCA AGGAGGTCAAT GGAGGGGGTG TATCCCCCTGG CCGTGCCCCT GGAGGTGGAG	2460
GTGGGGATAG GGGAGGACTG GCTCTCCGCC AAGGAGTGAT ACCACC	2506

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2496 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGGCGATGC TTCCCTCTT TGAGCCAAA GGCCGGTGC TCCTGGTGGA CGGCCACAC	60
CTGGCCTACC GCACCTCTT TGCCCTCAAG GGCTCACCA CGAGCCGCGG CGAACCCGTT	120
CAGGCGGTCT ACGGCTCGC CAAAAGCCTC CTCAAGGCCG TGAAGGAGGA CGGGGACGTG	180
GTGGTGGTGG TCTTGACGC CAAGGCCCCC TCCTTCCGCC ACGAGGCCTA CGAGGCCTAC	240
AAGGCGGGCC GGGCCCCCAC CCCGGAGGAC TTTCCCGGCG AGCTGGCCCT CATCAAGGAG	300
TTGGTGGACC TCCTAGGCCT TGTGCGGCTG GAGGTTCCCG GCTTGAGGC GGACGACGTG	360
CTGGCCACCC TGGCCAAGCG GGCGGAAAAG GAGGGTACG AGGTGCGCAT CCTCACTGCC	420
GACCGCGACC TCTACCAGCT CCTTCCGGAG CGCATGCCA TCCTCCACCC TGAGGGTAC	480
CTGATCACCC CGCGTGGCT TTACGAGAAG TACGGCCTGC GCCCGGAGCA GTGGGTGGAC	540
TACCGGGCCC TGGCGGGGGA CCCCTCGGAT AACATCCCCG GGGTGAAGGG CATCGGGAG	600
AAGACCGCCC AGAGGCTCAT CCGCGAGTGG GGGAGCCTGG AAAACCTCTT CCAGCACCTG	660
GACCAGGTGA AGCCCTCCTT GCGGGAGAAG CTCCAGGCCG GCATGGAGGC CCTGCCCTT	720
TCCCGGAAGC TTTCCCAGGT GCACACTGAC CTGCCCTGG AGGTGGACTT CGGGAGGCAG	780

CGCACACCCA	ACCTGGAGGG	TCTGCGGGCT	TTTTGGAGC	GGTTGGAGTT	TGGAAGCCTC	840
CTCCACGAGT	TCGGCCTCCT	GGAGGGGCCG	AAGGCAGCAG	AGGAGGCCCC	CTGGCCCCCT	900
CCGGAAGGGG	CTTTTTGGG	CTTTCCCTT	TCCC GTCCCG	AGCCCATGTG	GGCGAGCTT	960
CTGGCCCTGG	CTGGGGCGTG	GGAGGGCGC	CTCCATCGGG	CACAAGACCC	CCTTAGGGC	1020
CTGAGGGACC	TTAAGGGGT	GCGGGAAATC	CTGGCCAAGG	ACCTGGCGGT	TTTGGCCCTG	1080
CGGGAGGGCC	TGGACCTCTT	CCCAGAGGAC	GACCCCATGC	TCCTGGCCTA	CCTTCTGGAC	1140
CCCTCCAACA	CCACCCCTGA	GGGGGTGGCC	CGGCGTTACG	GGGGGGAGTG	GACGGAGGAT	1200
GCGGGGGAGA	GGGCCCTCCT	GGCCGAGCGC	CTCTTCCAGA	CCCTAAAGGA	GCGCCTTAAG	1260
GGAGAAGAAC	GCCTGCTTTG	GCTT TACGAG	GAGGTGGAGA	AGCCGCTTT	CCGGGTGTTG	1320
GCCCGGATGG	AGGCCACGGG	GGTCCGGCTG	GACGTGGCCT	ACCTCCAGGC	CCTCTCCCTG	1380
GAGGTGGAGG	CGGAGGTGCG	CCAGCTGGAG	GAGGAGGTCT	TCCGCCTGGC	CGGCCACCC	1440
TTCAACCTCA	ACTCCCGCGA	CCAGCTGGAG	CGGGTGCTCT	TTGACGAGCT	GGGCCTGCCT	1500
GCCATCGGCA	AGACGGAGAA	GACGGGGAAA	CGCTCCACCA	GCGCTGCCGT	GCTGGAGGCC	1560
CTGCGAGAGG	CCCACCCAT	CGTGGACCGC	ATCCTGCAGT	ACCGGGAGCT	CACCAAGCTC	1620
AAGAACACCT	ACATAGACCC	CCTGCCCGCC	CTGGTCCACC	CCAAGACCGG	CCGGCTCCAC	1680
ACCCGTTCA	ACCAGACGGC	CACCGCCACG	GGCAGGCTTT	CCAGCTCCGA	CCCCAACCTG	1740
CAGAACATCC	CCGTGCCAC	CCCTCTGGGC	CAGCGCATCC	GCCGAGCCTT	CGTGGCGAG	1800
GAGGGCTGGG	TGCTGGTGGT	CTTGGACTAC	AGCCAGATTG	AGCTTCGGGT	CCTGGCCCAC	1860
CTCTCCGGGG	ACGAGAACCT	GATCCGGTC	TTTCAGGAGG	GGAGGGACAT	CCACACCCAG	1920
ACCGCCAGCT	GGATGTCGG	CGTTTCCCCC	GAAGGGTAG	ACCCTCTGAT	GCGCCGGCG	1980
GCCAAGACCA	TCAACTTCGG	GGT GCTCTAC	GGCATGTCCG	CCCACCGCCT	CTCCGGGGAG	2040
CTTTCCATCC	CCTACGAGGA	GGCGGTGGCC	TTCATTGAGC	GCTACTTCCA	GAGCTACCCC	2100
AAGGTGCGGG	CCTGGATTGA	GGGGACCCCTC	GAGGAGGGCC	GCCGGCGGGG	GTATGTGGAG	2160
ACCCCTTCG	GCCGCCGGCG	CTATGTGCC	GACCTCAAACG	CCC GG GTGAA	GAGCGTGC	2220
GAGGC CGGCGG	AGCGCATGGC	CTTCAACATG	CCGGTCCAGG	GCACCGCCGC	CGACCTCATG	2280
AAGCTGGCCA	TGGTGGGCT	TTTCCCCCGG	CTTCAGGAAC	TGGGGCGAG	GATGCTTTG	2340
CAGGTGCACG	ACGAGCTGGT	CCTCGAGGCC	CCCAAGGACC	GGGCAGGAG	GGTAGCCGCT	2400
TTGGCCAAGG	AGGTCACTGGA	GGGGGTCTGG	CCCCTGCAGG	TGCCCCTGGA	GGTGGAGGTG	2460
GGCCTGGGGG	AGGACTGGCT	CTCCGCCAAG	GAGTAG			2496

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGAGGCGA	TGCTTCCGCT	CTTGAAACCC	AAAGGCCGGG	TCCTCCTGGT	GGACGCCAC	60
CACCTGGCCT	ACCGCACCTT	CTTCGCCCTG	AAGGGCCTCA	CCACGAGCCG	GGCGAACCG	120
GTGCAGGCGG	TCTACGGCTT	CGCCAAGAGC	CTCCTCAAGG	CCCTGAAGGA	GGACGGGTAC	180
AAGGCCGTCT	TCGTGGTCTT	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGAG	240
GCCTACAAGG	CGGGGAGGGC	CCCGACCCCC	GAGGACTTCC	CCCGGCAGCT	CGCCCTCATC	300
AAGGAGCTGG	TGGACCTCCT	GGGGTTTACC	CGCCTCGAGG	TCCCCGGCTA	CGAGGCGGAC	360
GACGTTCTCG	CCACCCCTGGC	CAAGAAGGCG	GAAAAGGAGG	GGTACGAGGT	GCGCATCCTC	420
ACCGCCGACC	GCGACCTCTA	CCAACTCGTC	TCCGACCGCG	TCGCCGTCTT	CCACCCCGAG	480
GGCCACCTCA	TCACCCCGGA	GTGGCTTGG	GAGAAGTACG	GCCTCAGGCC	GGAGCAGTGG	540
GTGGACTTCC	GCGCCCTCGT	GGGGGACCCC	TCCGACAACC	TCCCCGGGGT	CAAGGGCATC	600
GGGGAGAAGA	CCGCCCTCAA	GCTCCTCAAG	GAGTGGGAA	GCCTGGAAAA	CCTCCTCAAG	660
AACCTGGACC	GGGTAAAGCC	AGAAAACGTC	CGGGAGAAGA	TCAAGGCCCA	CCTGGAAGAC	720
CTCAGGCTCT	CCTTGGAGCT	CTCCCGGGTG	CGCACCGACC	TCCCCCTGGA	GGTGGACCTC	780
GCCCAAGGGC	GGGAGCCCGA	CCGGGAGGGG	CTTAGGGCCT	TCCTGGAGAG	GCTGGAGTTC	840
GGCAGCCTCC	TCCACGAGTT	CGGCCTCCTG	GAGGCCCGCG	CCCCCTGGA	GGAGGCCCCC	900
TGGCCCCCGC	CGGAAGGGC	CTTCGTGGGC	TTCGTCCCT	CCCGCCCCGA	GCCCATGTGG	960
GCGGAGCTTA	AAGCCCTGGC	CGCCTGCAGG	GACGGCCGGG	TGCACCGGGC	AGCAGACCCC	1020
TTGGCGGGGC	TAAAGGACCT	CAAGGAGGTC	CGGGGCCTCC	TCGCCAAGGA	CCTCGCCGTC	1080
TTGGCCTCGA	GGGAGGGCT	AGACCTCGTG	CCCGGGGACG	ACCCCATGCT	CCTCGCCTAC	1140
CTCCTGGACC	CCTCCAACAC	CACCCCGAG	GGGGTGGCGC	GGCGCTACGG	GGGGGAGTGG	1200
ACGGAGGACG	CCGCCAACCG	GGCCCTCCTC	TCGGAGAGGC	TCCATCGGAA	CCTCCTTAAG	1260
CGCCTCGAGG	GGGAGGAGAA	GCTCCTTTGG	CTCTACCACG	AGGTGGAAA	GCCCTCTCC	1320
CGGGTCCTGG	CCCACATGGA	GGCCACCGGG	GTACGGCTGG	ACGTGGCCTA	CCTTCAGGCC	1380
CTTCCCTGG	AGCTTGGGA	GGAGATCCGC	CGCCTCGAGG	AGGAGGTCTT	CCGCTTGGCG	1440
GGCCACCCCT	TCAACCTCAA	CTCCCGGGAC	CAGCTGGAAA	GGGTGCTCTT	TGACGAGCTT	1500

... A C T G T C G A T C C C G C C G G T G ...

AGGCTTCCCG CCTTGGGAA GACGAAAAG ACAGGCAAGC GCTCCACCAG CGCCGCGGTG 1560
 CTGGAGGCC TACGGGAGGC CCACCCCCATC GTGGAGAAGA TCCTCCAGCA CGGGAGCTC 1620
 ACCAACGCTCA AGAACACCTA CGTGGACCCC CTCCAAGCC TCGTCCACCC GAGGACGGC 1680
 CGCCTCCACA CCCGCTCAA CCAGACGGCC ACGGCCACGG GGAGGCTTAG TAGCTCCGAC 1740
 CCCAACCTGC AGAACATCCC CGTCCGCACC CCCTTGGGCC AGAGGATCCG CGGGCCTTC 1800
 GTGGCCGAGG CGGGTTGGGC GTTGGTGGCC CTGGACTATA GCCAGATAGA GCTCCCGTC 1860
 CTCGCCCACC TCTCCGGGA CGAAAACCTG ATCAGGGTCT TCCAGGAGGG GAAGGACATC 1920
 CACACCCAGA CCGCAAGCTG GATGTTCGGC GTCCCCCGG AGGCCGTGGA CCCCTGATG 1980
 CGCCGGCGG CCAAGACGGT GAACCTCGGC GTCCTCTACG GCATGTCCGC CCATAGGCTC 2040
 TCCCAGGAGC TTGCCATCCC CTACGAGGAG GCGGTGGCCT TTATAGAGGC TACTTCCAAA 2100
 GCTTCCCCAA GGTGCGGGCC TGGATAGAAA AGACCCTGGA GGAGGGGAGG AAGCGGGCT 2160
 ACGTGGAAAC CCTCTTCGGA AGAAGGCGCT ACGTGCCCGA CCTCAACGCC CGGGTGAAGA 2220
 GCGTCAGGGA GGCGCGGAG CGCATGGCCT TCAACATGCC CGTCCAGGGC ACCGCCGCG 2280
 ACCTCATGAA GCTCGCCATG GTGAAGCTCT TCCCCCGCCT CGGGGAGATG GGGGCCGCA 2340
 TGCTCCTCCA GGTCCACGAC GAGCTCCTCC TGGAGGCCCG CCAAGCGCGG GCCGAGGAGG 2400
 TGGCGGCTTT GGCAAGGAG GCCATGGAGA AGGCCTATCC CCTCGCCGTG CCCCTGGAGG 2460
 TGGAGGTGGG GATGGGGAG GACTGGCTTT CCGCCAAGGG TTAG 2504

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 832 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Arg	Gly	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu
1				5				10					15		
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	His	Ala	Leu	Lys	Gly
				20				25					30		
Leu	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala
				35				40				45			
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Asp	Ala	Val	Ile	Val
				50				55				60			
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Gly	Gly
				65				70				75			80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
 85 90 95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu
 100 105 110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys
 115 120 125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp
 130 135 140

Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly
 145 150 155 160

Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
 165 170 175

Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn
 180 185 190

Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu
 195 200 205

Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu
 210 215 220

Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys
 225 230 235 240

Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val
 245 250 255

Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe
 260 265 270

Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
 275 280 285

Glu Ser Pro Lys Ala Leu Glu Ala Pro Trp Pro Pro Pro Glu Gly
 290 295 300

Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp
 305 310 315 320

Leu Leu Ala Leu Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro
 325 330 335

Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu
 340 345 350

Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro
 355 360 365

Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
 370 375 380

Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
 385 390 395 400

Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu
 405 410 415

Trp Gly Arg Leu Glu Gly Glu Arg Leu Leu Trp Leu Tyr Arg Glu
420 425 430
Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly
435 440 445
Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala
450 455 460
Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His
465 470 475 480
Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
485 490 495
Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
500 505 510
Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
515 520 525
Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr
530 535 540
Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu
545 550 555 560
His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
565 570 575
Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
580 585 590
Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala
595 600 605
Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly
610 615 620
Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
625 630 635 640
Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro
645 650 655
Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
660 665 670
Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu
675 680 685
Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg
690 695 700
Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val
705 710 715 720
Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg
725 730 735
Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro
740 745 750

Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
 755 760 765
 Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His
 770 775 780
 Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala
 785 790 795 800
 Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
 805 810 815
 Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 820 825 830

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 831 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Val
 1 5 10 15
 Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
 20 25 30
 Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
 35 40 45
 Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Val Val Val Val
 50 55 60
 Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala Tyr
 65 70 75 80
 Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu Ala
 85 90 95
 Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Val Arg Leu Glu Val
 100 105 110
 Pro Gly Phe Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Arg Ala
 115 120 125
 Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Asp Leu
 130 135 140
 Tyr Gln Leu Leu Ser Glu Arg Ile Ala Ile Leu His Pro Glu Gly Tyr
 145 150 155 160
 Leu Ile Thr Pro Ala Trp Leu Tyr Glu Lys Tyr Gly Leu Arg Pro Glu
 165 170 175
 Gln Trp Val Asp Tyr Arg Ala Leu Ala Gly Asp Pro Ser Asp Asn Ile
 180 185 190

Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Gln Arg Leu Ile Arg
 195 200 205
 Glu Trp Gly Ser Leu Glu Asn Leu Phe Gln His Leu Asp Gln Val Lys
 210 215 220
 Pro Ser Leu Arg Glu Lys Leu Gln Ala Gly Met Glu Ala Leu Ala Leu
 225 230 235 240
 Ser Arg Lys Leu Ser Gln Val His Thr Asp Leu Pro Leu Glu Val Asp
 245 250 255
 Phe Gly Arg Arg Arg Thr Pro Asn Leu Glu Gly Leu Arg Ala Phe Leu
 260 265 270
 Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu Glu
 275 280 285
 Gly Pro Lys Ala Ala Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly Ala
 290 295 300
 Phe Leu Gly Phe Ser Phe Ser Arg Pro Glu Pro Met Trp Ala Glu Leu
 305 310 315 320
 Leu Ala Leu Ala Gly Ala Trp Glu Gly Arg Leu His Arg Ala Gln Asp
 325 330 335
 Pro Leu Arg Gly Leu Arg Asp Leu Lys Gly Val Arg Gly Ile Leu Ala
 340 345 350
 Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Phe Pro
 355 360 365
 Glu Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr
 370 375 380
 Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu Asp
 385 390 395 400
 Ala Gly Glu Arg Ala Leu Leu Ala Glu Arg Leu Phe Gln Thr Leu Lys
 405 410 415
 Glu Arg Leu Lys Gly Glu Glu Arg Leu Leu Trp Leu Tyr Glu Glu Val
 420 425 430
 Glu Lys Pro Leu Ser Arg Val Leu Ala Arg Met Glu Ala Thr Gly Val
 435 440 445
 Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Glu Ala
 450 455 460
 Glu Val Arg Gln Leu Glu Glu Val Phe Arg Leu Ala Gly His Pro
 465 470 475 480
 Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu
 485 490 495
 Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser
 500 505 510
 Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val
 515 520 525

Asp Arg Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Asn Thr Tyr
 530 535 540
 Ile Asp Pro Leu Pro Ala Leu Val His Pro Lys Thr Gly Arg Leu His
 545 550 555 560
 Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser
 565 570 575
 Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg
 580 585 590
 Ile Arg Arg Ala Phe Val Ala Glu Glu Gly Trp Val Leu Val Val Leu
 595 600 605
 Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp
 610 615 620
 Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Gln
 625 630 635 640
 Thr Ala Ser Trp Met Phe Gly Val Ser Pro Glu Gly Val Asp Pro Leu
 645 650 655
 Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met
 660 665 670
 Ser Ala His Arg Leu Ser Gly Glu Leu Ser Ile Pro Tyr Glu Glu Ala
 675 680 685
 Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Tyr Pro Lys Val Arg Ala
 690 695 700
 Trp Ile Glu Gly Thr Leu Glu Glu Gly Arg Arg Gly Tyr Val Glu
 705 710 715 720
 Thr Leu Phe Gly Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg Val
 725 730 735
 Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val
 740 745 750
 Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Arg Leu Phe
 755 760 765
 Pro Arg Leu Gln Glu Leu Gly Ala Arg Met Leu Leu Gln Val His Asp
 770 775 780
 Glu Leu Val Leu Glu Ala Pro Lys Asp Arg Ala Glu Arg Val Ala Ala
 785 790 795 800
 Leu Ala Lys Glu Val Met Glu Gly Val Trp Pro Leu Gln Val Pro Leu
 805 810 815
 Glu Val Glu Val Gly Leu Gly Glu Asp Trp Leu Ser Ala Lys Glu
 820 825 830

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 834 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe
50 55 60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
65 70 75 80

Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
85 90 95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu
100 105 110

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys
115 120 125

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg
130 135 140

Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
145 150 155 160

Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg
165 170 175

Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp
180 185 190

Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu
195 200 205

Leu Lys Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg
210 215 220

Val Lys Pro Glu Asn Val Arg Glu Lys Ile Lys Ala His Leu Glu Asp
225 230 235 240

Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp Leu Pro Leu
245 250 255

Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu Gly Leu Arg
260 265 270

Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly
 275 280 285
 Leu Leu Glu Ala Pro Ala Pro Leu Glu Ala Pro Trp Pro Pro Pro
 290 295 300
 Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp
 305 310 315 320
 Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg
 325 330 335
 Ala Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly
 340 345 350
 Leu Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp
 355 360 365
 Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro
 370 375 380
 Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp
 385 390 395 400
 Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg
 405 410 415
 Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr
 420 425 430
 His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala
 435 440 445
 Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu
 450 455 460
 Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Val Phe Arg Leu Ala
 465 470 475 480
 Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu
 485 490 495
 Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly
 500 505 510
 Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His
 515 520 525
 Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys
 530 535 540
 Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly
 545 550 555 560
 Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu
 565 570 575
 Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu
 580 585 590
 Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu
 595 600 605

Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu
 610 615 620
 Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile
 625 630 635 640
 His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val
 645 650 655
 Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu
 660 665 670
 Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr
 675 680 685
 Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys
 690 695 700
 Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly
 705 710 715 720
 Tyr Val Glu Thr Leu Phe Gly Arg Arg Tyr Val Pro Asp Leu Asn
 725 730 735
 Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn
 740 745 750
 Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val
 755 760 765
 Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln
 770 775 780
 Val His Asp Glu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu
 785 790 795 800
 Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala
 805 810 815
 Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala
 820 825 830
 Lys Gly

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2502 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGNNGGCGA TGCTTCCCCT CTTTGAGCCC AAAGGCCGGG TCCTCCTGGT GGACGGCCAC	60
CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA CCACCAGCCG GGGCGAACCG	120
GTGCAGGCAG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTGAAGGA GGACGGGGAC	180

NNGCGGGTGN	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGAG	240
GCCTACAAGG	CGGGCCGGGC	CCCCACCCCG	GAGGACTTTC	CCCGGCAGCT	300
AAGGAGCTGG	TGGACCTCCT	GGGGCTTGCG	CGCCTCGAGG	TCCCCGGCTA	360
GACGTNCTGG	CCACCCCTGGC	CAAGAAGGCG	GAAAAGGAGG	GGTACGAGGT	420
ACCGCCGACC	GCGACCTCTA	CCAGCTCCTT	TCCGACCGCA	TCGCCGTCT	480
GGGTACCTCA	TCACCCCCGGC	GTGGCTTTGG	GAGAAGTACG	GCCTGAGGCC	540
GTGGACTACC	GGGCCCTGGC	GGGGGACCCC	TCCGACAACC	TCCCCGGGGT	600
GGGGAGAAGA	CCGCCCNGAA	GCTCCTCNAG	GAGTGGGGGA	GCCTGGAAAA	660
AACCTGGACC	GGGTGAAGCC	CGCCNTCCGG	GAGAAGATCC	AGGCCACAT	720
ANGCTCTCCT	GGGAGCTNTC	CCAGGTGCGC	ACCGACCTGC	CCCTGGAGGT	780
AAGNGCGGG	AGCCCGACCG	GGAGGGCTT	AGGGCCTTTC	TGGAGAGGCT	840
AGCCTCTCC	ACGAGTCGG	CCTCCTGGAG	GGCCCCAAGG	CCCTGGAGGA	900
CCCCCGCCGG	AAGGGCCTT	CGTGGGCTTT	GTCCTTCCC	GCCCCGAGCC	960
GAGCTCTGG	CCCTGGCCGC	CGCCAGGGAG	GGCCGGGTCC	ACCGGGCACCC	1020
ANGGGCCTNA	GGGACCTNA	GGAGGTGCGG	GGNCTCCTCG	CCAAGGACCT	1080
GCCCTGAGGG	AGGGCCTNGA	CCTCNTGCC	GGGGACGACC	CCATGCTCCT	1140
CTGGACCCCT	CCAACACCAAC	CCCCGAGGGG	GTGGCCCGGC	GCTACGGGGG	1200
GAGGANGCGG	GGGAGCGGGC	CCTCCTNTCC	GAGAGGCTCT	TCCNGAACCT	1260
CTTGAGGGGG	AGGAGAGGCT	CCTTGGCTT	TACCAGGAGG	TGGAGAAGCC	1320
GTCCTGGCCC	ACATGGAGGC	CACGGGGTN	CGGCTGGACG	TGGCCTACCT	1380
TCCCTGGAGG	TGGCGGAGGA	GATCCGCCGC	CTCGAGGAGG	AGGTCTTCCG	1440
CACCCCTTCA	ACCTCAACTC	CCGGGACCAG	CTGGAAAGGG	TGCTCTTGA	1500
CTTCCCGCCA	TCGGCAAGAC	GGAGAAGACN	GGCAAGCGCT	CCACCAGCGC	1560
GAGGCCCTNC	GNGAGGCCA	CCCCATCGTG	GAGAAGATCC	TGCAGTACCG	1620
AAGCTCAAGA	ACACCTACAT	NGACCCCCCTG	CCNGNCCTCG	TCCACCCAG	1680
CTCCACACCC	GCTTCAACCA	GACGGCCACG	GCCACGGCA	GGCTTAGTAG	1740
AACCTGCAGA	ACATCCCCGT	CCGCACCCCN	CTGGGCCAGA	GGATCCGCCG	1800
GCCGAGGAGG	GNTGGGTGTT	GGTGGCCCTG	GACTATAGCC	AGATAGAGCT	1860
GCCCACCTCT	CCGGGGACGA	GAACCTGATC	CGGGTCTTCC	AGGAGGGAG	1920
ACCCAGACCG	CCAGCTGGAT	GTTCGGCGTC	CCCCCGGAGG	CCGTGGACCC	1980
CGGGCGGCCA	AGACCATCAA	CTTCGGGGTC	CTCTACGGCA	TGTCCGCCA	2040

CAGGAGCTTG CCATCCCTA CGAGGAGGCG GTGGCCTTCA TTGAGCGCTA CTTCCAGAGC 2100
 TTCCCCAAGG TGCAGGGCTG GATTGAGAAG ACCCTGGAGG AGGGCAGGAG GCAGGGGTAC 2160
 GTGGAGACCC TCTTCGGCCG CCGCGCTAC GTGCCGACC TCAACGCCG GGTGAAGAGC 2220
 GTGCGGGAGG CGGCGGAGCG CATGGCCTTC AACATGCCG TCCAGGGCAC CGCCGCCGAC 2280
 CTCATGAAGC TGGCCATGGT GAAGCTCTTC CCCCGGCTNC AGGAAATGGG GGCCAGGATG 2340
 CTCCTNCAGG TCCACGACGA GCTGGTCCTC GAGGCCCCA AAGAGCGGGC GGAGGNGGTG 2400
 GCCGCTTGG CCAAGGAGGT CATGGAGGGG GTCTATCCCC TGGCCGTGCC CCTGGAGGTG 2460
 GAGGTGGGGA TGGGGGAGGA CTGGCTCTCC GCCAAGGAGT AG 2502

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 833 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Xaa	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu
1				5				10					15		
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly
	20						25						30		
Leu	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala
	35					40						45			
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Asp	Ala	Val	Xaa	Val
	50					55					60				
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	Ala
	65					70			75				80		
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
	85						90					95			
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Leu	Xaa	Arg	Leu	Glu
	100					105						110			
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
	115					120					125				
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp
	130					135					140				
Leu	Tyr	Gln	Leu	Leu	Ser	Asp	Arg	Ile	Ala	Val	Leu	His	Pro	Glu	Gly
	145					150					155		160		
Tyr	Leu	Ile	Thr	Pro	Ala	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
	165						170					175			

Glu Gln Trp Val Asp Tyr Arg Ala Leu Xaa Gly Asp Pro Ser Asp Asn
 180 185 190
 Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Xaa Lys Leu Leu
 195 200 205
 Xaa Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg Val
 210 215 220
 Lys Pro Xaa Xaa Arg Glu Lys Ile Xaa Ala His Met Glu Asp Leu Xaa
 225 230 235 240
 Leu Ser Xaa Xaa Leu Ser Xaa Val Arg Thr Asp Leu Pro Leu Glu Val
 245 250 255
 Asp Phe Ala Xaa Arg Arg Glu Pro Asp Arg Glu Gly Leu Arg Ala Phe
 260 265 270
 Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
 275 280 285
 Glu Xaa Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly
 290 295 300
 Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp Ala Glu
 305 310 315 320
 Leu Leu Ala Leu Ala Ala Arg Xaa Gly Arg Val His Arg Ala Xaa
 325 330 335
 Asp Pro Leu Xaa Gly Leu Arg Asp Leu Lys Glu Val Arg Gly Leu Leu
 340 345 350
 Ala Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Xaa
 355 360 365
 Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
 370 375 380
 Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
 385 390 395 400
 Asp Ala Gly Glu Arg Ala Leu Leu Ser Glu Arg Leu Phe Xaa Asn Leu
 405 410 415
 Xaa Xaa Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Xaa Glu
 420 425 430
 Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr Gly
 435 440 445
 Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Ala
 450 455 460
 Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His
 465 470 475 480
 Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
 485 490 495
 Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
 500 505 510

Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
 515 520 525
 Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Asn Thr
 530 535 540
 Tyr Ile Asp Pro Leu Pro Xaa Leu Val His Pro Arg Thr Gly Arg Leu
 545 550 555 560
 His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
 565 570 575
 Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
 580 585 590
 Arg Ile Arg Arg Ala Phe Val Ala Glu Glu Gly Trp Xaa Leu Val Ala
 595 600 605
 Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly
 610 615 620
 Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
 625 630 635 640
 Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp Pro
 645 650 655
 Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
 660 665 670
 Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu
 675 680 685
 Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg
 690 695 700
 Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val
 705 710 715 720
 Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg
 725 730 735
 Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro
 740 745 750
 Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
 755 760 765
 Phe Pro Arg Leu Xaa Glu Met Gly Ala Arg Met Leu Leu Gln Val His
 770 775 780
 Asp Glu Leu Val Leu Glu Ala Pro Lys Xaa Arg Ala Glu Xaa Val Ala
 785 790 795 800
 Ala Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
 805 810 815
 Leu Glu Val Glu Val Gly Xaa Gly Glu Asp Trp Leu Ser Ala Lys Glu
 820 825 830
 Xaa

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1647 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGAATTCTGG GGATGCTGCC CCTCTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC	60
CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CGGGGGGGAG	120
CCGGTGCAGG CGGTCTACGG CTCGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG	180
GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG	240
GGGTACAAGG CGGGCCGGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC	300
AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC	360
GACGTCCCTGG CCAGCCTGGC CAAGAACGGC GAAAAGGAGG GCTACGAGGT CGCGATCCTC	420
ACCGCCGACA AAGACCTTA CCAGCTCCTT TCCGACCGCA TCCACGTCTT CCACCCCGAG	480
GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG	540
GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCCTC	600
GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG	660
AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAACATCC TGGCCCACAT GGACGATCTG	720
AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC	780
AAAAGGCGGG AGCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTGGC	840
AGCCTCCTCC ACGAGTCGG CCTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG	900
CCCCCGCCGG AAGGGCCCTT CGTGGCTTT GTGCTTCCC GCAAGGAGCC CATGTGGGCC	960
GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCC CGACCTTAT	1020
AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG	1080
GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GGCGACGACC CCATGCTCCT CGCCTACCTC	1140
CTGGACCTT CCAACACCAC CCCCGAGGGG GTGGCCGGC GCTACGGCGG GGAGTGGACG	1200
GAGGAGGCGG GGGAGCGGGC CGCCCTTCC GAGAGGCTCT TCGCCAACCT GTGGGGAGG	1260
CTTGAGGGGG AGGAGAGGCT CCTTCTGGTT TACCGGGAGG TGGAGAGGCC CCTTCCGCT	1320
GTCCTGGCCC ACATGGAGGC CACGGGGGTG CGCCTGGACG TGGCCTATCT CAGGGCCTTG	1380
TCCCTGGAGG TGGCCGGGA GATCGCCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC	1440
CACCCCTTCA ACCTCAACTC CCGGGACCAAG CTGGAAAGGG TCCTCTTGA CGAGCTAGGG	1500

CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCTG	1560
GAGGCCCTCC GCGAGGCCA CCCCATCGTG GAGAAGATCC TGCAGGCATG CAAGCTTGGC	1620
ACTGGCCGTC GTTTACAAC GTCGTGA	1647

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2088 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGAATTCTGG GGATGCTGCC CCTCTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC	60
CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAAG CCGGGGGGAG	120
CCGGTGCAGG CGGTCTACGG CTTGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG	180
GACGCCGTGA TCGTGGCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG	240
GGGTACAAGG CGGGCCGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC	300
AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC	360
GACGTCTGG CCAGCCTGGC CAAGAAGGGC GAAAAGGAGG GCTACGAGGT CGCGATCCTC	420
ACCGCCGACA AAGACCTTA CCAGCTCCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG	480
GGGTACCTCA TCACCCCGC CTGGCTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG	540
GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC	600
GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG	660
AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG	720
AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC	780
AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTGGC	840
AGCCTCCTCC ACGAGTTCGG CCTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG	900
CCCCCGCCGG AAGGGCCTT CGTGGGCTTT GTGCTTCCC GCAAGGAGCC CATGTGGGCC	960
GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCG CGAGCCTTAT	1020
AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG	1080
GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GGCGACGACC CCATGTCCT CGCCTACCTC	1140
CTGGACCCCTT CCAACACCAC CCCCGAGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG	1200
GAGGAGGCCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG	1260
CTTGAGGGGG AGGAGAGGCT CCTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTCCGCT	1320

GTCTGGCCC ACATGGAGGC CACGGGGTG CGCCTGGACG TGGCTATCT CAGGGCCTTG	1380
TCCCTGGAGG TGGCCGGGA GATCGCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC	1440
CACCCCTTCA ACCTCAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTGA CGAGCTAGGG	1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG	1560
GAGGCCCTCC GCGAGGCCA CCCCATCGT GAGAAGATCC TGCACTACCG GGAGCTCAC	1620
AAGCTGAAGA GCACCTACAT TGACCCCTTG CCGGACCTCA TCCACCCAG GACGGCCGC	1680
CTCCACACCC GCTTCAACCA GACGGCCACG GCCACGGCA GGCTAAGTAG CTCCGATCCC	1740
AACCTCCAGA ACATCCCCGT CCGCACCCCG CTTGGCAGA GGATCCGCCG GGCCTTCATC	1800
GCCGAGGAGG GGTGGCTATT GGTGGCCCTG GACTATAGCC AGATAGAGCT CAGGGTGCTG	1860
GCCCACCTCT CCGGCACGA AACCTGATC CGGGTCTTCC AGGAGGGCGG GGACATCCAC	1920
ACGGAGACCG CCAGCTGGAT GTTCGGCGTC CCCCGGGAGG CCGTGGACCC CCTGATGCGC	1980
CGGGCGGCCA AGACCATCAA CTTCGGGTC CTCTACGGCA TGTCGGCCCA CGCCTCTCC	2040
CAGGAGCTAG CTAGCCATCC CTTACGAGGA GGCCAGGCC TTCATTGA	2088

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 962 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGAATTCTGG GGATGCTGCC CCTCTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC	60
CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCAACAG CGGGGGGGAG	120
CCGGTGCAGG CGGTCTACGG CTTGCCAAG AGCCTCTCA AGGCCCTCAA GGAGGACGGG	180
GACGCGGTGA TCGTGGCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG	240
GGGTACAAGG CGGGCCGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC	300
AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCAGGCTA CGAGGCGGAC	360
GACGTCTGG CCAGCCTGGC CAAGAACCGC GAAAAGGAGG GCTACGAGGT CCGCATCCTC	420
ACCGCCGACA AAGACCTTA CCAGCTTCTT TCCGACCGCA TCCACGTCCT CCACCCGAG	480
GGGTACCTCA TCACCCCGC CTGGCTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG	540
GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC	600
GGGGAGAAGA CGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG	660
AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG	720

AAGCTCTCCT	GGGACCTGGC	CAAGGTGC	ACCGACCTGC	CCCTGGAGGT	GGACTTCGCC	780
AAAAGGCGGG	AGCCCGACCG	GGAGAGGCTT	AGGGCCTTTC	TGGAGAGGCT	TGAGTTGGC	840
AGCCTCCTCC	ACGAGTCGG	CCTCTGGAA	AGCCCCAAGT	CATGGAGGGG	GTGTATCCCC	900
TGGCCGTGCC	CCTGGAGGTG	GAGGTGGGA	TAGGGGAGGA	CTGGCTCTCC	GCCAAGGAGT	960
GA						962

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1600 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGAATT	CG GGGATGCTGC	CCCTCTTG	GA CCCAAGGGC	CGGGTCCTCC	TGGTGGACGG	60
CCACCACCTG	GCCTACCGCA	CCTTCCACGC	CCTGAAGGGC	CTCACCA	GCCGGGGGGA	120
GCCGGTGCAG	GCGGTCTACG	GCTTCGCCAA	GAGCCTCCTC	AAGGCCCTCA	AGGAGGACGG	180
GGACGCGGTG	ATCGTGGTCT	TTGACGCCAA	GGCCCCCTCC	TTCCGCCACG	AGGCCTACGG	240
GGGGTACAAG	GCGGGCCGGG	CCCCCACGCC	GGAGGACTTT	CCCCGGCAAC	TCGCCCTCAT	300
CAAGGAGCTG	GTGGACCTCC	TGGGCTGGC	GCGCCTCGAG	GTCCCGGGCT	ACGAGGCGGA	360
CGACGTCTG	GCCAGCCTGG	CCAAGAACGC	GGAAAAGGAG	GGCTACGAGG	TCCGCATCCT	420
CACCGCCGAC	AAAGACCTTT	ACCAGCTCCT	TTCCGACC	ATCCACGTCC	TCCACCCC	480
GGGGTACCTC	ATCACCCGG	CCTGGCTTTG	GGAAAAGTAC	GGCCTGAGGC	CCGACCAGTG	540
GGCCGACTAC	CGGGCCCTGA	CCGGGGACGA	GTCCGACAAC	CTTCCC	TCAAGGGCAT	600
CGGGGAGAAG	ACGGCGAGGA	AGCTTCTGGA	GGAGTGGGGG	AGCCTGGAAG	CCCTCCTCAA	660
GAACCTGGAC	CGGCTGAAGC	CCGCCATCCG	GGAGAACATC	CTGGCCCACA	TGGACGATCT	720
GAAGCTCTCC	TGGGACCTGG	CCAAGGTGCG	CACCGACCTG	CCCCTGGAGG	TGGACTTCGC	780
CAAAAGGCGG	GAGCCCGACC	GGGAGAGGCT	TAGGGCCTTT	CTGGAGAGGC	TTGAGTTGG	840
CAGCCTCCTC	CACGAGTTCG	GCCTCTGGA	AAGCCCCAAG	ATCCGCC	GGGG CCGGG	900
CGAGGAGGGG	TGGCTATTGG	TGGCCCTGGA	CTATAGCCAG	ATAGAGCTCA	GGGTGCTGGC	960
CCACCTCTCC	GGCGACGAGA	ACCTGATCCG	GGTCTCCAG	GAGGGGCGGG	ACATCCACAC	1020
GGAGACCGCC	AGCTGGATGT	TCGGCGTCCC	CCGGGAGGCC	GTGGACCCCC	TGATGCGCCG	1080
GGCGGCCAAG	ACCATCAACT	TCGGGGTCCT	CTACGGCATG	TCGGCCCACC	GCCTCTCCCA	1140
GGAGCTAGCC	ATCCCTTACG	AGGAGGCCA	GGCCTTCATT	GAGCGCTACT	TTCAAGAGCTT	1200

CCCCAAGGTG CGGGCCTGGA TTGAGAAGAC CCTGGAGGAG GGCAGGAGGC GGGGGTACGT	1260
GGAGACCCCTC TTCGGCGGCC GCCGCTACGT GCCAGACCTA GAGGCCCGGG TGAAGAGCGT	1320
GCAGGAGGCG GCCGAGCGCA TGGCCTCAA CATGCCGTC CGGGGCACCG CCGCCGACCT	1380
CATGAAGCTG GCTATGGTGA AGCTCTTCCC CAGGCTGGAG GAAATGGGGG CCAGGATGCT	1440
CCTTCAGGTC CACGACCGAGC TGTCCTCGA GGCCCCAAA GAGAGGGCGG AGGCCGTGGC	1500
CCGGCTGGCC AAGGAGGTCA TGGAGGGGGT GTATCCCCTG GCCGTGCCCG TGGAGGTGGA	1560
GGTGGGGATA GGGGAGGACT GGCTCTCCGC CAAGGAGTGA	1600

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CACGAATTTCG GGGATGCTGC CCCTCTTGA GCCCAA	36
---	----

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGAGATCTA TCACTCCTTG GCGGAGAGCC AGTC	34
---------------------------------------	----

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 91 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TAATACGACT CACTATAGGG AGACCGGAAT TCGAGCTCGC CCGGGCGAGC TCGAATTCCG	60
---	----

TGTATTCTAT AGTGTACCT AAATCGAATT C	91
-----------------------------------	----

)
})

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TAATACGACT CACTATAGGG

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GAATTCGATT TAGGTGACAC TATAGAA

27

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTAATCATGG TCATAGCTGG TAGCTTGCTA C

31

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGATCCTCTA GAGTCGACCT GCAGGCATGC CTACCTTGGT AG

42

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGATCCTCTA GAGTCGACCT GCAGGCATGC 30

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2502 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGAATTCCG	GGATGCTGCC	CCTCTTGAG	CCCAAGGGCC	GGGTCCCTCCT	GGTGGACGGC	60
CACCACCTGG	CCTACCGCAC	CTTCCACGCC	CTGAAGGGCC	TCACCACCAAG	CCGGGGGGAG	120
CCGGTGCAGG	CGGTCTACGG	CTTCGCCAAG	AGCCTCCTCA	AGGCCCTCAA	GGAGGACGGG	180
GACGCGGTGA	TCGTGGTCTT	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGGG	240
GGGTACAAGG	CGGGCCGGGC	CCCCACGCCG	GAGGACTTTC	CCCGGCAACT	CGCCCTCATC	300
AAGGAGCTGG	TGGACCTCCT	GGGGCTGGCG	CGCCTCGAGG	TCCCAGGCTA	CGAGGCAGGAC	360
GACGTCTGG	CCAGCCTGGC	CAAGAACCG	GAAAAGGAGG	GCTACGAGGT	CCGCATCCTC	420
ACCGCCGACA	AAGACCTTTA	CCAGCTCCTT	TCCGACCGCA	TCCACGTCCT	CCACCCCGAG	480
GGGTACCTCA	TCACCCCGGC	CTGGCTTGG	GAAAAGTACG	GCCTGAGGCC	CGACCAGTGG	540
GCCGACTACC	GGGCCCTGAC	CGGGGACGAG	TCCGACAACC	TTCCCGGGGT	CAAGGGCCTC	600
GGGGAGAAGA	CGGCGAGGAA	GCTTCTGGAG	GAGTGGGGGA	GCCTGGAAGC	CCTCCTCAAG	660
AACCTGGACC	GGCTGAAGCC	CGCCATCCGG	GAGAAGATCC	TGGCCCACAT	GGACGATCTG	720
AAGCTCTCCT	GGGACCTGGC	CAAGGTGCGC	ACCGACCTGC	CCCTGGAGGT	GGACTTCGCC	780
AAAAGGCAGG	AGCCCGACCG	GGAGAGGCTT	AGGGCCTTTC	TGGAGAGGCT	TGAGTTGGC	840
AGCCTCCTCC	ACGAGTTCGG	CCTTCTGGAA	AGCCCCAAGG	CCCTGGAGGA	GGCCCCCTGG	900
CCCCCGCCGG	AAGGGCCTT	CGTGGGCTTT	GTGCTTCCC	GCAAGGAGCC	CATGTGGGCC	960
GATCTTCTGG	CCCTGGCCGC	CGCCAGGGGG	GGCCGGGTCC	ACCGGGCCCC	CGAGCCTTAT	1020
AAAGCCCTCA	GGGACCTGAA	GGAGGCGCGG	GGGCTTCTCG	CCAAAGACCT	GAGCGTTCTG	1080

GCCCTGAGGG	AAGGCCTTGG	CCTCCCGCCC	GGCGACGACC	CCATGCTCCT	CGCCTACCTC	1140
CTGGACCCCTT	CCAACACCAAC	CCCCGAGGGG	GTGGCCCGGC	GCTACGGCGG	GGAGTGGACG	1200
GAGGAGGCAGG	GGGAGCGGGC	CGCCCTTCC	GAGAGGCTCT	TCGCCAACCT	GTGGGGGAGG	1260
CTTGAGGGGG	AGGAGAGGCT	CCTTTGGCTT	TACCGGGAGG	TGGAGAGGCC	CCTTTCCGCT	1320
GTCCCTGGCCC	ACATGGAGGC	CACGGGGGTG	CGCCTGGACG	TGGCCTATCT	CAGGGCCTTG	1380
TCCCTGGAGG	TGGCCGGGA	GATCGCCCGC	CTCGAGGCCG	AGGTCTTCCG	CCTGGCCGGC	1440
CACCCCTTCA	ACCTCAACTC	CCGGGACCAAG	CTGGAAAGGG	TCCTCTTGA	CGAGCTAGGG	1500
CTTCCCGCCA	TCGGCAAGAC	GGAGAAAGACC	GGCAAGCGCT	CCACCAGCGC	CGCCGTCCTG	1560
GAGGCCCTCC	GCGAGGCCCA	CCCCATCGTG	GAGAAGATCC	TGCAGTACCG	GGAGCTCAC	1620
AAGCTGAAGA	GCACCTACAT	TGACCCCTTG	CCGGACCTCA	TCCACCCAG	GACGGGCCGC	1680
CTCCACACCC	GCTTCAACCA	GACGGCCACG	GCCACGGCA	GGCTAAGTAG	CTCCGATCCC	1740
AACCTCCAGA	ACATCCCCGT	CCGCACCCCG	CTTGGCAGA	GGATCCGCCG	GGCCTTCATC	1800
GCCGAGGAGG	GGTGGCTATT	GGTGGCCCTG	GACTATAGCC	AGATAGAGCT	CAGGGTGCTG	1860
GCCCACCTCT	CCGGCGACGA	GAACCTGATC	CGGGCTTCC	AGGAGGGCG	GGACATCCAC	1920
ACGGAGACCG	CCAGCTGGAT	GTTCGCGTC	CCCCGGGAGG	CCGTGGACCC	CCTGATGCGC	1980
CGGGCGGCCA	AGACCATCAA	CTTCGGGGTC	CTCTACGGCA	TGTCGGCCCA	CCGCCTCTCC	2040
CAGGAGCTAG	CCATCCCTTA	CGAGGAGGCC	CAGGCCTCA	TTGAGCGCTA	CTTCAGAGC	2100
TTCCCCAAGG	TGCGGGCTG	GATTGAGAAG	ACCCCTGGAGG	AGGGCAGGAG	GCGGGGTAC	2160
GTGGAGACCC	TCTTCGGCCG	CCGCCGCTAC	GTGCCAGACC	TAGAGGCCCG	GGTGAAGAGC	2220
GTGCGGGAGG	CGGCCGAGCG	CATGGCCTTC	AACATGCCG	TCCGGGGCAC	CGCCGCCGAC	2280
CTCATGAAGC	TGGCTATGGT	GAAGCTCTTC	CCCAGGCTGG	AGGAAATGGG	GGCCAGGATG	2340
CTCCTTCAGG	TCCACGACGA	GCTGGTCCTC	GAGGCCCAA	AAGAGAGGGC	GGAGGCCGTG	2400
GCCCGGCTGG	CCAAGGAGGT	CATGGAGGGG	GTGTATCCCC	TGGCGTGCC	CCTGGAGGTG	2460
GAGGTGGGGA	TAGGGGAGGA	CTGGCTCTCC	GCCAAGGAGT	GA		2502

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GATTTAGGTG ACACTATAG

19

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGGACGAACA AGCGAGACAG CGACACAGGT ACCACATGGT ACAAGAGGCA AGAGAGACGA	60
CACAGCAGAA AC	72

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTTTCTGCTG TGTCGTCTCT CTTGCCTCTT GTACCATGTG GTACCTGTGT CGCTGTCTCG	60
CTTGTTCGTC	70

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GACGAACAAG CGAGACAGCG	20
-----------------------	----

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTTTCTGCTG TGTCGTCTCT CTTG	24
----------------------------	----

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCTCTTGTAC CATGTGGTAC CTGTGTCGCT GTCTCGCTTG TTTCGTC

46

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACACAGGTAC CACATGGTAC AAGAGGCAAG AGAGACGACA CAGCAGAAAC

50

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Asn Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 969 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCA ATTCGGGGAT GCTGCCCTC

60

TTTGAGCCA AGGGCCGGGT CCTCCTGGTG GACGGCCACC ACCTGGCCTA CCGCACCTTC

120

CACGCCCTGA AGGGCCTCAC CACCAGCCGG GGGGAGCCGG TGCAGGCGGT CTACGGCTTC

180

GCCAAGAGCC TCCTCAAGGC CCTCAAGGAG GACGGGGACG CGGTGATCGT GGTCTTTGAC

240

GCCAAGGCC	CCTCCTCCG	CCACGAGGCC	TACGGGGGT	ACAAGGCAGG	CCGGGCCCCC	300
ACGCCGGAGG	ACTTTCCCCG	GCAACTCGCC	CTCATCAAGG	AGCTGGTGGA	CCTCCTGGGG	360
CTGGCGCGCC	TCGAGGTCCC	GGGCTACGAG	GCGGACGACG	TCCTGGCCAG	CCTGGCCAAG	420
AAGGCGGAAA	AGGAGGGCTA	CGAGGTCCGC	ATCCTCACCG	CCGACAAAGA	CCTTTACCAG	480
CTTCTTCCG	ACCGCATCCA	CGTCCTCCAC	CCCGAGGGGT	ACCTCATCAC	CCCGGCCTGG	540
CTTTGGGAAA	AGTACGGCCT	GAGGCCGAC	CAGTGGGCCG	ACTACCGGGC	CCTGACCGGG	600
GACGAGTCCG	ACAACCTTCC	CGGGGTCAAG	GGCATCGGGG	AGAAGACGGC	GAGGAAGCTT	660
CTGGAGGAGT	GGGGGAGCCT	GGAAGCCCTC	CTCAAGAAC	TGGACCGGGC	GAAGCCCGCC	720
ATCCGGGAGA	AGATCCTGGC	CCACATGGAC	GATCTGAAGC	TCTCCTGGGA	CCTGGCCAAG	780
GTGCGCACCG	ACCTGCCCT	GGAGGTGGAC	TTCGCCAAAA	GGCGGGAGCC	CGACCGGGAG	840
AGGCTTAGGG	CCTTCTGGA	GAGGCTTGAG	TTTGGCAGCC	TCCTCCACGA	GTTCGGCCTT	900
CTGGAAAGCC	CCAAGTCATG	GAGGGGGTGT	ATCCCCTGGC	CGTGCCCTG	GAGGTGGAGG	960
TGGGGATAG						969

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 948 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCA	ATTGGGGAT	GCTGCCCTC	60
TTTGAGCCA	AGGGCCGGGT	CCTCCTGGTG	GACGGCCACC	ACCTGGCCTA	CCGCACCTTC	120
CACGCCCTGA	AGGGCCTCAC	CACCAGCCGG	GGGGAGCCGG	TGCAGGCGGT	CTACGGCTTC	180
GCCAAGAGCC	TCCTCAAGGC	CCTCAAGGAG	GACGGGGACG	CGGTGATCGT	GGTCTTGAC	240
GCCAAGGCC	CCTCCTCCG	CCACGAGGCC	TACGGGGGT	ACAAGGCAGG	CCGGGCCCCC	300
ACGCCGGAGG	ACTTTCCCCG	GCAACTCGCC	CTCATCAAGG	AGCTGGTGGA	CCTCCTGGGG	360
CTGGCGCGCC	TCGAGGTCCC	GGGCTACGAG	GCGGACGACG	TCCTGGCCAG	CCTGGCCAAG	420
AAGGCGGAAA	AGGAGGGCTA	CGAGGTCCGC	ATCCTCACCG	CCGACAAAGA	CCTTTACCAG	480
CTTCTTCCG	ACCGCATCCA	CGTCCTCCAC	CCCGAGGGGT	ACCTCATCAC	CCCGGCCTGG	540
CTTTGGGAAA	AGTACGGCCT	GAGGCCGAC	CAGTGGGCCG	ACTACCGGGC	CCTGACCGGG	600
GACGAGTCCG	ACAACCTTCC	CGGGGTCAAG	GGCATCGGGG	AGAAGACGGC	GAGGAAGCTT	660
CTGGAGGAGT	GGGGGAGCCT	GGAAGCCCTC	CTCAAGAAC	TGGACCGGGC	GAAGCCCGCC	720

ATCCGGGAGA AGATCCTGGC CCACATGGAC GATCTGAAGC TCTCCTGGGA CCTGGCCAAG GTGCGCACCG ACCTGCCCT GGAGGTGGAC TTCGCCAAA GGCAGGGAGCC CGACCGGGAG AGGCTTAGGG CCTTTCTGGA GAGGCTTGAG TTTGGCAGCC TCCTCCACGA GTTCGGCCTT CTGGAAAGCC CCAAGGCCGC ACTCGAGCAC CACCACCAAC ACCACTGA	780 840 900 948
(2) INFORMATION FOR SEQ ID NO:32:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 206 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
CGCCAGGGTT TTCCCAGTCA CGACGTTGTA AAACGACGGC CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCTC TAGAGTCGAC CTGCAGGCAT GCAAGCTTGA GTATTCTATA GTGTCACCTA AATAGCTTGG CGTAATCATG GTCATAGCTG TTTCCCTGTGT GAAATTGTTA TCCGCT	60 120 180 206
(2) INFORMATION FOR SEQ ID NO:33:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
TTCTGGGTTC TCTGCTCTCT GGTGCTGTC TCGCTTGTTC GTC	43
(2) INFORMATION FOR SEQ ID NO:34:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
GCTGTCTCGC TTGTTGTC	19

{ (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GACGAACAAG CGAGACAGCG

20

{ (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTCTGGGTTC TCTGCTCTCT GGTG

24

{ (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GACGAACAAG CGAGACAGCG ACCAGAGAGC AGAGAACCCCA GAA

43

{ (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ACCAGAGAGC AGAGAACCCCA GAA

23

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AACAGCTATG ACCATGATTA C

21

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTTCTCTGCT CTCTGGTCGC TGTCTCGCTT GTGAAACAAG CGAGACAGCG TGTTCTCTCG

60

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGAGAGACCA CGCTG

15

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTGCCCG GCTTTCCCCG TC

52

()

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AGAAAGGAAG GGAAGAAAGC GAAAGG

26

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GACGGGGAAA GCCGGCGAAC G

21

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAAAGCCGGC GAACGTGGCG

20

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GGCGAACGTG GCGAGAAAGG A

21

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CCTTTCGCTT TCTTCCCTTC CTTTCTGCC ACGTTGCCCG GC

42

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CCTTTCGCTC TCTTCCCTTC CTTTCTGCC ACGTTGCCCG GC

42

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 8
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

/mod_base= OTHER

/note= "The A residue at this position is 2'-O-methyladenosine."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AGAAAGGAAG GGAAGAAAGC GAAAGGT

27

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

<p>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:</p> <p>GCCGGCGAAC GTGGCGAGAA AGGA</p> <p>(2) INFORMATION FOR SEQ ID NO:51:</p> <p>(i) SEQUENCE CHARACTERISTICS:</p> <ul style="list-style-type: none"> (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear <p>(ii) MOLECULE TYPE: DNA (genomic)</p> <p>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:</p> <p>GGTTTTTCTT TGAGGTTTAG</p> <p>(2) INFORMATION FOR SEQ ID NO:52:</p> <p>(i) SEQUENCE CHARACTERISTICS:</p> <ul style="list-style-type: none"> (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear <p>(ii) MOLECULE TYPE: DNA (genomic)</p> <p>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:</p> <p>GCGACACTCC ACCATAGAT</p> <p>(2) INFORMATION FOR SEQ ID NO:53:</p> <p>(i) SEQUENCE CHARACTERISTICS:</p> <ul style="list-style-type: none"> (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear <p>(ii) MOLECULE TYPE: DNA (genomic)</p> <p>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:</p> <p>CTGTCTTCAC GCAGAAAGC</p> <p>(2) INFORMATION FOR SEQ ID NO:54:</p> <p>(i) SEQUENCE CHARACTERISTICS:</p> <ul style="list-style-type: none"> (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear <p>(ii) MOLECULE TYPE: DNA (genomic)</p> <p>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:</p> <p>GCACGGTCTA CGAGACCTC</p>	<p>24</p> <p>20</p> <p>19</p> <p>19</p> <p>19</p>
--	---

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TAATACGACT CACTATAGGG

20

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGAAAGCUU GCAUGCUGC AGGUCGACUC UAGAGGAUCU ACUAGUCAUA UGGAUUCUGU	60
CUUCACGCAG AAAGCGUCUG GCCAUGGCGU UAGUAUGAGU GUCGUGCAGC CUCCAGGACC	120
CCCCCUCCCCG GGAGAGGCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCAGG	180
ACGACCGGGU CCUUUCUUGG AUAAACCCGC UCAAUGCCUG GAGAUUUGGG CGUGCCCCCG	240
CAAGACUGCU AGCCGAGUAG UGUUGGGUCG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG	300
UGCCUGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGC	337

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 18
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

/note= "The N at this position indicates the presence of a
fluorescein dye on an abasic linker."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCGGTCGTCC TGGCAATNCC

20

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GTTTATCCAA GAAAGGACCC GGTCC

25

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CAGGGTGAAG GGAAGAAGAA AGCGAAAGGT

30

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CAGGGGGAAG GGAAGAAGAA AGCGAAAGGT

30

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..2
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

/mod_base= OTHER

/note= "The T residues at positions 1 and 2 are amino modified T residues."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TTCTTTCAC CAGCGAGACG GG

22

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

ATTGGGCGCC AGGGTGGTTT TT

22

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 53 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CCCGTCTCGC TGGTGAAAG AAAAAACCACCC CTGGCGCCCA ATACGCAAAC CGC

53

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAATTCGATT TAGGTGACAC TATAGAATAC A

31

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CCTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTGCCCG GC

42

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GCCCCGCGAAC GTGGCGAGAA AGGA

24

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CAGAAAGGAAG GGAAGAAAGC GAAAGG

26

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CAGGGGGAAG GGAAGAAAGC GAAAGG

26

(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CAGGGTACAG GGAAGAAAGC GAAAGG

26